Proposition of convolutional neural network based system for skin cancer detection

Esther CHABI ADJOBO[1, 2*], Amadou Tidjani SANDA MAHAMA[1, 2], Pierre GOUTON[2], Joël TOSSA[1]
1 Institute of Mathematics and Physics Sciences, BP 613, Porto-Novo, Benin;
2 University of Burgundy, BP 47870, 21078 Dijon Cedex, France;
*Corresponding author: esther.chabi@imsp-uac.org

***Abstract*—Skin cancer automated diagnosis tools play a vital role in timely screening, helping dermatologists focus on melanoma cases. Best arts on automated melanoma screening use deep learning-based approaches, especially deep convolutional neural networks (CNN) to improve performances. Because of the large number of parameters that could be involved during training in CNN many training samples are needed to avoid overfitting problem. Gabor filtering can efficiently extract spatial information including edges and textures, which may reduce the features extraction burden to CNN. In this paper, we proposed a Gabor Convolutional Network (GCN) model to improve the performance of automated diagnosis of skin cancer systems. The model combines a CNN model and Gabor filtering and serves three functions: generation of Gabor filter banks, CNN construction and filter injection. We performed experiments with dermoscopic images and results were interpreted according to classification accuracy. The results we have obtained show that our GCN offers the best classification accuracy with a value of 96.39% against 94.02% for the CNN model.**

Keywords— Skin cancer, CNN, GCN, Gabor filtering, dermoscopic images.

I. INTRODUCTION

The human skin is the outer covering of the body and is the largest organ of the integumentary system. Skin has mesodermal cells, pigmentation, such as melanin provided by melanocytes, which absorb some of the potentially dangerous ultraviolet radiation (UV) in sunlight. It also contains DNA repair enzymes that help reverse UV damage, such that people lacking the genes for these enzymes and people overexposed to UV suffer high rates of skin cancer. There are three major types of skin cancers: basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and melanoma. The first two skin cancers are grouped together as non-melanoma skin cancers [1]. The incidence of both non-melanoma and melanoma skin cancers has been increasing over the past decades. Every year in America, non-melanoma skin cancer affects more than 3 million people. Melanoma rates in the United States doubled from 1982 to 2011 and have continued to increase [2] .

Skin cancer including melanoma and non-melanoma is highly curable if diagnosed in its early stage (Fig.1). However, early diagnosis is very challenging as melanomas are easily confused with benign skin lesions (Fig.2).

The first and most common procedure performed daily by dermatologists is visual evaluation of lesions. In this context, dermatologists rely primarily on aspects of a skin lesion to determine whether it is benign or malignant. Visual evaluation is based on the distribution, size, shape, border, symmetry and colour of lesion. Diagnosis based on colour are subjective as colour perception depends on the human visual response to the light and light interaction with skin [3].

To improve diagnosis, dermoscopy method also termed epiluminescence microscopy (ELM) was introduced. Dermoscopy has opened a new dimension in the examination of pigmented skin lesions and, especially, in the identification of the early phase of cutaneous malignant melanoma [4]. Dermoscopy is a non-invasive method that allows the in vivo evaluation of colours and microstructures of the epidermis, the dermoepidermal junction, and the papillary dermis that are not visible to the naked eye [5]. It has also been revealed that dermoscopy in the hands of inexperienced dermatologists may cause a reduction in diagnostic accuracy. This implies a need to develop more reliable and robust system for the diagnosis of skin cancer.

Efforts have been done to develop Computer Assisted Diagnostic (CAD) for melanoma diagnose based on skin lesion images. CAD systems use artificial intelligence to analyze lesion data and reach a diagnosis of skin cancer. Automated tools help experts and non-expert focus on patients or at-risk injuries. CAD systems can be based on traditional Machine Learning techniques (classical image processing techniques) or Convolutional Neural Networks (CNNs). Classical image processing techniques consist of image pre-processing (hair, and occlusions removal), image segmentation, features extractions (hand-crafted local and global features) and classification (non-linear classification algorithms) [6]. With CNNs, no need to extract manually features, the network takes directly the whole picture as input, learns features automatically [7] and performs classification task.

In this paper, we work on a GCN (Gabor-CNN) dermoscopic image analysis system that aids dermatologists to more appropriately triage high-risk lesions. The objective of this study is to apply new trends of CNNs on dermoscopic images to accurately classify skin lesions.

The paper is organized as follows: a review of previous related works is presented in section 1, the proposed model for skin cancer classification in Section 2. Into Section 3, experimental results and discussion are reported in detail and finally Section 4 draws the conclusion.

978-1-7281-5686-6/19/$31.00 ©2019 IEEE
DOI 10.1109/SITIS.2019.00018

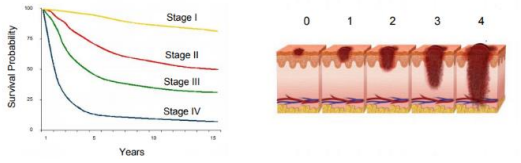

Figure 1: Skin cancer stages and survival probability [1]

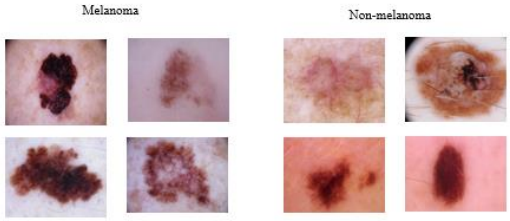

Figure 2: Sample images created from ISIC 2019 dataset

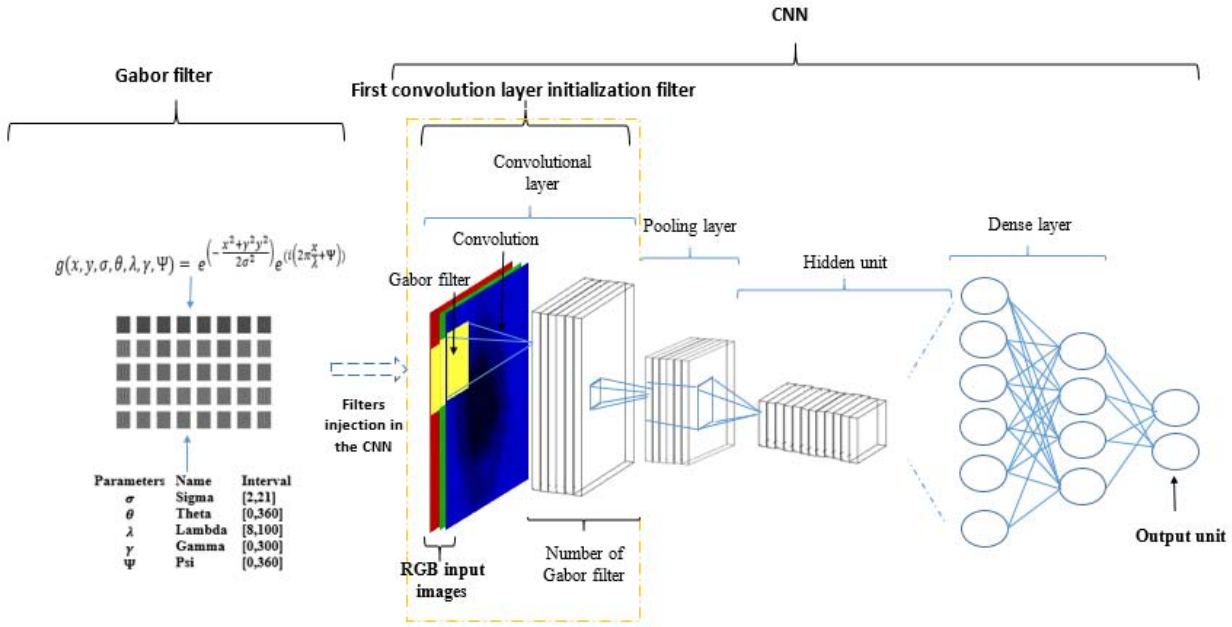

Figure 3: Framework of the Gabor filtering and CNN (GCN)

II. RELATED WORKS

There is a vast literature on automated screening for melanoma but until 2015, the literature is mostly based on classical image processing techniques [8]. With the success of Deep Learning techniques in Computer Vision, melanoma screening research are more and more based on Deep Learning models.

CNNs are used to classify skin lesions in two fundamentally different ways.

On the one hand, a network pre-trained on another large dataset, such as ImageNet [9], can be used as a feature extractor. In this case, classification is performed by another classifier, such as k-nearest neighbours, support vector machines, or artificial neural Networks.

Reference [10] used a dataset of 399 images, collected from DermIS [11]. The data was classified into two classes: benign nevi and malignant melanoma. Features were automatically extracted from the last three layer of the CNN named AlexNet and k nearest neighbours classifier (kNN) was used. The performance of the classifier is calculated in terms of sensitivity (%) =85.71±2.1, specificity (%) =82.48±6.8 and accuracy (%) = 83.95±3.3. Reference [12] proposed an approach that combines deep learning, sparse coding, and support vector machine (SVM) learning algorithms for melanoma recognition in dermoscopic images. The dataset was obtained from the International Skin Imaging Collaboration, containing 2,624 images. The approach is compared to the prior state-of-art method on the same dataset. Two-fold cross-validation was performed 20 times for evaluation (40 total experiments), and

TABLE 1 : PROPOSED MODEL SUMMURY

Layer (type)	*Output Shape*	*Parameters*
conv2d (Conv2D)	(None, 224, 224, 96)	34944
avg_pooling2d (MaxPooling2D)	(None, 112, 112, 96)	0
conv2d_1 (Conv2D)	(None, 112, 112, 256)	614656
batch_normalization (BatchNo	(None, 112, 112, 256)	1024
max_pooling2d_1 (MaxPooling2	(None, 56, 56, 256)	0
batch_normalization_1 (Batch	(None, 56, 56, 256)	1024
conv2d_2 (Conv2D)	(None, 56, 56, 384)	885120
conv2d_3 (Conv2D)	(None, 56, 56, 384)	1327488
conv2d_4 (Conv2D)	(None, 56, 56, 256)	884992
max_pooling2d_2 (MaxPooling2	(None, 28, 28, 256)	0
flatten (Flatten)	(None, 200704)	0
dense (Dense)	(None, 512)	102760960
dense_1 (Dense)	(None, 1)	513

two discrimination tasks are examined: 1) melanoma vs. all non-melanoma lesions, and 2) melanoma vs. atypical lesions only. The presented approach achieves an accuracy of 93.1% (94.9% sensitivity, and 92.8% specificity) for the first task, and 73.9% accuracy (73.8% sensitivity, and 74.3% specificity) for the second task.

On the other hand, a CNN can directly learn the relationship between the raw pixel data and the class labels through end-to-end learning. In contrast with the classical workflow typically applied in machine learning, feature extraction becomes an integral part of classification and is no longer considered as a separate, independent processing step. If the CNN is trained by end-to-end learning, the research can be additionally divided into two different approaches: learning the model from scratch or transfer learning [13].

Reference [14] have proposed a learning from scratch method based on Deep Residual Network. The dataset contain 3,600 images from ISIC 2017. The training was performed by fine-tuned the ResNet model, pre-trained on ImageNet. Averaged area under curve (AUC) was used for evaluation of the three classification approaches (1) multi-class classification AUC= 90.60; (2) binary classification approach AUC= 91.30; and (3) the ensemble approach AUC= 91.50.

Reference [15] used the most extensive dataset of 129,450 clinical images, including 3,374 dermoscopy images. They used transfer learning from a pre-trained Inception v3 model, removed the decision layer and added a new decision layer with 757 classes independent set of diseases. After fine tuning the model, they compared the specificity and sensitivity of over 20 human specialists with their model's, concluding that the deep architecture employed outperforms the specialists, on average.

Researcher's results show that deeper models outperform shallower models [8], [10] and [15] but there is no literature

TABLE 2 : ACCURACY OF GCN AND CNN ON HALF DATASET

	Train		*Test*	
Epoch	*GCN*	*CNN*	*GCN*	*CNN*
1	0.7525	0.6564	0.6432	0.5113
2	0.8041	0.768	0.8047	0.8359
3	0.9102	0.8931	0.8828	0.5182
4	0.8911	0.8936	0.8724	0.849
5	0.9199	0.9052	0.7969	0.7734
6	0.9162	0.9162	0.9062	0.8776
7	0.9272	0.9118	0.8385	0. 7109
8	0.9173	0.9019	0.9115	0.9001
9	0.9118	0.9184	0.9219	0. 8802
10	0.9609	0.9035	0.8906	0.8698
11	0.9449	0.9283	0.9036	0.9193
12	0.9443	0.9328	0.9062	0.8776
13	0.9639	0.9052	0.9479	0.875
14	0.9544	0.9512	0.8984	0.8229
15	0.9702	0.9305	0.9479	0.8672

TABLE 3 : ACCURACY OF GCN AND CNN ON FULL DATASET

	Train		*Test*	
Epoch	*GCN*	*CNN*	*GCN*	*CNN*
1	0.8045	0.6698	0.6787	0.5996
2	0.9122	0.8624	0.7900	0.6074
3	0.9102	0.8833	0.9053	0.6289
4	0.9106	0.9016	0.7607	0.6582
5	0.9209	0.9048	0.9160	0.7471
6	0.9178	0.9117	0.8936	0.7305
7	0.9454	0.9255	0.9062	0.8027
8	0.9570	0.9372	0.9395	0.9043
9	0.9566	0.9199	0.8926	0.8672
10	0.9540	0.9471	0.9502	0.8984
11	0.9515	0.9351	0.9590	0.9258
12	0.9741	0.9663	0.9473	0.9072
13	0.9898	0.9395	0.9531	0.9180
14	0.9614	0.9699	0.9551	0.9365
15	0.9811	0.9571	0.9639	0.9402

indicating the depth required for an accurate diagnostic deep network for skin cancer.

CNNs are more and more essential in images classification, but their use for medical images is difficult because these models require very large training sets (tens of thousands to millions of images) [16]. To get around this difficulty, current literature employs transfert learning, a technique where a model formed for a source task is partially "recycled" for a new target task. According to [13], the source data type affect the learning accuracy obtained from the new network.

The combination of CNN and Gabors filter bank is a new method in image classification with deep learning. Named GCNN (Gabor CNN) or GCN (Gabor Convolutional Network)

or CNN gabor-type filters, this method consists of initializing CNN filters with a Gabor filters bank. [17] and [18] build GCN for general classification task on MNIST, CIFAR10 and CIFAR100 datasets and by experiments, they observed accuracy perfomance, training time reduction and optimization of memory use. Reference [19] used a GCN to enhance the resistance of deep learned features to the orientation and scale changes in object recognition task. Their experimental results demonstrated GCN's ability to recognize objects when changes in scale and rotation occur frequently. GCN are also used in [20] for object recognition in natural scene. The GCN were used for strengthening the learning of texture information. Through experiments, their approach achieved the recognition rate of 81.53%, yielding a 1.26% promotion in the average accuracy rate compared with the results obtained using the convolutional neural network model alone on the ImageNet10 dataset.

In the next section, we proposed an adapted GCN model for accurate skin cancer classification.

III. PROPOSED GCN MODEL

The framework of our proposed method is shown in Fig. 3. The first step is the generation of Gabor filters bank. Gabor filter bank is a set of linear filter constructed to respond the edges and textures of varying frequencies and orientations. Such a designed Gabor filter bank exhibits a vision system similar to human visual perception [21]. Hence, Gabor filters are successful for classification tasks which are usual for human vision. Gabor function formula is as follows:

$$g(x, y, \sigma, \theta, \lambda, \gamma, \Psi) = e^{\left(-\frac{x^2+\gamma^2 y^2}{2\sigma^2}\right)} e^{(i\left(2\pi\frac{x}{\lambda}+\Psi\right))} \quad (1)$$

where σ is the standard deviation of the Gaussian function, θ is the filter orientation, λ is the wavelength of the sinusoidal factor, γ is the spatial aspect factor and Ψ is the phase offset [17].

On the second step, we build the model. The model architecture is presented in Table 1. This model consists of five convolution blocks activated by a *ReLU* activation function, one average pooling layer, two max pooling layers, two batch normalization layers and one fully connected layer. The model output is activated by the sigmoid activation function. The strengths of the model are: overlap pooling to reduce the size of network and batch normalization to accelerate deep network training by reducing internal covariate shift.

The third step consist of initialization of the first convolution layer of the model with Gabor filters initialization. It is used to extract robust and discriminant features for the subsequent step. The first layer of the model is the most important layer in the architecture as we changed the initialization of this layer from the default Xavier uniform distribution to our Gabor filters initialization. During training the layers of the model and parameters are updated through optimization algorithm (Stochastic Gradient Descent) and back propagated through loss function (Categorical Cross Entropy).

IV. EXPERIMENTAL RESULTS

Data description

For the experiments we use a dermoscopic dataset to validate the performance of the proposed method. The dataset is available for free on the International Skin Imaging Collaboration (ISIC 2019) image archive [22]. The image archive contained 33,569 dermoscopic images organized into 9 classes: actinic Keratosis, basal Cell Carcinoma, benign Keratosis, dermato fibroma, melanocytic Nevus, melanoma, squamous Cell Carcinoma, vascular Lesion [22] [23] [24]. We reorganized the dataset into 2 classes: melanoma with 4,523 images and non-melanoma (all dataset images except melanoma images) with 20,785 and remove images whose outlines are not visible. 80% of images were used for training and 20% for test.

Results and discussions

We performed experiments on Gabor initialized models (GCN) and default initialized models (CNN) in 15 epochs. The models were compiled with the ADAM optimizer and a learning rate of 0.01.

The Table 2 detail the performance of the GCN and CNN models. The dataset has been split in two.

In Table 3, the performance of the GCN and CNN models are listed in detail. The models were compiled with the complete dataset.

In both tables, Gabor initialized model have about 9.61% (Table 2) and 13.47% (Table 3) improvement in early stages of the training. On test set, Gabor initialized model have a significant improvement against the CNN model and this gap continues in later phase of training. In the end, both model almost converges with CNN model achieve good performance but, by injected Gabor filters, the accuracy value increase.

Literature on [17] and [25] proof that texture features contains relevant information for melanoma and non-melanoma classification. So by using Gabor filtering, important features are captured. Obtained results confirm that, on the one hand, texture analysis is important to distinguish melanomas from non-melanomas. On the other hand, the CNN model perform the classification task. According to literature [14], [13], [24] and [15], deep learning methods in general perform the classification task but the accuracy depend on the training set [17]. This work shows that Gabor's filtering can alleviate the CNN dependence on the size of the training dataset.

V. CONCLUSION

In this paper, we propose a method that combines Gabor filtering and CNN model for skin cancer diagnosis. CNN is a powerful feature extraction method, but it requires a large number of learning samples to avoid overfitting. To reduce CNN overfitting, Gabor filtering has been used to efficiently extract spatial information, including edges and textures. We generate

a Gabor convolutional filter bank and used it to replace the random filter cores in the first convolutional layer. In order to evaluate the usefulness and performance of proposed model, experimentations are performed on a dermoscopic dataset ISIC 2019. The results obtained by the proposed method outperforms the create CNN model results even in case where the number of training samples is limited.

This study should be extended to various deep networks, multispectral images in order to test the generalization ability of the proposed method.

REFERENCES

[1] American Cancer Society, "Cancer Facts & Figures," 2017. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf. [Accessed 25 10 2019].

[2] H. Rogers, M. Weinstock, S. Feldman and B. Coldiron, "Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012," *JAMA Dermatol,* vol. 151, no. 10, pp. 1081-1087, 2015.

[3] R. Jolivot, "Development of an imaging system dedicated to the acquisition, analysis and multispectral characterisation of skin lesions," LABORATOIRE LE2I, Bourgogne, 2011.

[4] I. Stanganelli and M. A. Pizzichetta, "Dermoscopy," 13 Mars 2018. [Online]. Available: https://emedicine.medscape.com/article/1130783-overview.

[5] M. Zortea, "Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists," *Artificial Intelligence in Medicine,* p. 14, 2013.

[6] M. Fornaciali, M. Carvalho, F. V. Bittencourt, S. Avila and E. Valle, "Towards automated melanoma screening: Proper computer vision & reliable results," *arXiv:1604.04024v3,* pp. 1-15, 2016.

[7] MathWorks, "Introducing Deep Learning with MATLAB," 2018. [Online]. Available: https://www.mathworks.com/content/dam/mathworks/tag-team/Objects/d/80879v00_Deep_Learning_ebook.pdf. [Accessed 25 10 2019].

[8] P. Sameena, P. K. Gopalakrishna and S. PC, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review," *Biomedical Signal Processing and Control,* pp. 237-262, 2018.

[9] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *International Conference on Neural Information Processing Systems,* p. 1097–1105, 2012.

[10] V. Pomponiu, H. Nejati and C. N-M, "Deepmole: Deep neural networks for skin mole lesion classification," in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

[11] DermIS, "Dermatolody Information System," 04 09 2019. [Online]. Available: https://www.dermis.net/dermisroot/en/home/index.htm.

[12] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *MLMI: International Workshop on Machine Learning in Medical Imaging*, Munich, Germany, 2015.

[13] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk and C. v. Kalle, "Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review," *Journal of Medical Internet Research,* vol. 20, no. 10, pp. 1-8, 2018.

[14] L. Bi, J. Kim, E. Ahn and D. Feng, "Automatic Skin Lesion Analysis using Large-scale Dermoscopy Images and Deep Residual Networks," 04 09 2019. [Online]. Available: http://arxiv.org/pdf/1703.04197 webcite.

[15] A. Esteva and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, p. 115–118, 2017.

[16] H. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1285 - 1298, 2016.

[17] Ö. Gökhan and H. K. Ekenel, "Initialization of Convolutional Neural Networks by Gabor Filters," in *https://ieeexplore.ieee.org/document/8404757*, Izmir, Turkey, 2018.

[18] S. S. Sarwar, P. Panda and K. Roy, "Gabor Filter Assisted Energy Efficient Fast Learning Convolutional Neural Networks," in *2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, Taipei, Taiwan, 2017.

[19] L. Shangzhen, Z. Baochang, C. Chen, Z. Baochang, H. Jungong and L. Jianzhuang, "Gabor Convolutional Networks," *IEEE Transactions on Image Processing,* vol. 27, no. 9, pp. 1-11, 2018.

[20] Y. Hu, C. Li, D. Hu and W. Yu, "Gabor Feature Based Convolutional Neural Network for Object Recognition in Natural Scene," in *2016 3rd International Conference on Information Science and Control Engineering*, Beijing, China, 2016.

[21] M. Mermillod, P. Vuilleumier, C. Peyrin, D. Alleysson and C. Marendaz, "The importance of low spatial frequency information for recognising fearful facial expressions," *Connection Science,* vol. 21, no. 1, pp. 75-83, 2009.

[22] I. 2019, "ISIC 2019," [Online]. Available: https://challenge2019.isic-archive.com/. [Accessed 19 09 2019].

[23] P. Tschandl, C. Rosendahl and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data,* vol. 5, 2018.

[24] M. Combalia, N. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, A. C. Halpern, S. Puig and J. Malvehy, "BCN20000: Dermoscopic Lesions in the Wild," arXiv:1908.02288, Barcelona, 2019.

[25] A. Ebtihal and J. M. Arfan, "Classification of Dermoscopic Skin Cancer Images Using Color and Hybrid Texture Features," *IJCSNS International Journal of Computer Science and Network Security,,* vol. 6, no. 4, pp. 135-139, 2016.